Approximation Hardness for Small Occurrence Instances of NP-Hard Problems

Miroslav Chlebík[1] and Janka Chlebíková[2*]

[1] MPI for Mathematics in the Sciences, D-04103 Leipzig
[2] CAU, Institut für Informatik und Praktische Mathematik, D-24098 Kiel,
jch@informatik.uni-kiel.de

Abstract. The paper contributes to the systematic study (started by Berman and Karpinski) of explicit approximability lower bounds for small occurrence optimization problems. We present parametrized reductions for some packing and covering problems, including 3-Dimensional Matching, and prove the best known inapproximability results even for highly restricted versions of them. For example, we show that it is NP-hard to approximate MAX-3-DM within $\frac{139}{138}$ even on instances with exactly two occurrences of each element. Previous known hardness results for bounded occurence case of the problem required that the bound is at least three, and even then no explicit lower bound was known.
New structural results which improve the known bounds for 3-regular amplifiers and hence the inapproximability results for numerous small occurrence problems studied earlier by Berman and Karpinski are also presented.

1 Introduction

The research on the hardness of bounded occurrence (resp. bounded degree) optimization problems is focused on the case of very small value of the bound parameter. For many small parameter problems tight hardness results for optimization problems can be hardly achieved directly from the PCP characterization of NP. Rather, one has to use an expander/amplifier method. Considerable effort of Berman and Karpinski (see [2] and references therein) has gone into the developing of a new method of reductions for determining the inapproximability of MAXIMUM INDEPENDENT SET (MAX-IS) and MINIMUM NODE COVER (MIN-NC) in graphs of maximum degree 3 or 4.

Overview. As a starting point to our gap preserving reductions we state in Section 2 the versions of NP-hard gap results on bounded (constant) occurrence MAX-E3-LIN-2. Their weaker forms are known to experts and have been already used ([3], [7], [11], [12]). The advantage of this approach is that we need not restrict ourselves to amplifiers that can be constructed in polynomial time, to prove NP-hard gap results. Any (even nonconstructive) proof of existence of amplifiers (or expanders) with better parameters than those currently known implies the *existence* of (deterministic, polynomial) gap-preserving reductions

leading to better inapproximability result. This is our paradigm towards tighter inapproximability results inspired by the paper of Papadimitriou and Vempala on Traveling Salesman problem ([11]), that we have already used for Steiner Tree problem in [7].

We prove structural results about 3-regular amplifiers which play a crucial role in proving explicit inapproximability results for bounded occurrence optimization problems. A $(2,3)$-graph $G = (V, E)$ with nodes only of degree 2 (*Contacts*) and 3 (*Checkers*) is an *amplifier* if for very $A \subseteq V$ either $|\text{Cut } A| \geq |Contacts \cap A|$, or $|\text{Cut } A| \geq |Contacts \setminus A|$. The parameter $\tau(G) := \frac{|V|}{|Contacts|}$ measures the quality of an amplifier. We are able to prove for many bounded occurrence problems a tight correspondence between $\tau_* := \inf\{\gamma : \tau(G) < \gamma$ for infinity many amplifiers $G\}$ and inapproximability results. In this paper we slightly improve the upper bound from known $\tau_* \leq 7$ (Berman and Karpinski, [2]) to $\tau_* \leq 6.9$. This improvement is based on our structural amplifier analysis presented in Section 3. But there is still a substantial gap between the best upper and lower bounds on parameters of amplifiers and expanders. We develop our method of parametrized reductions (a parameter is a fixed amplifier) to prove inapproximability results for E3-Occ-Max-3-Lin-2 problem, and problems Max-IS and Min-NC on 3-regular graphs (Section 4). The similar method can be applied to all problems studied in [2] (with modification of amplifiers to bipartite-like for Max Cut) to improve the lower bound on approximability. Similarly, for the problem TSP with distances 1 and 2 ([3]).

We include reductions to some packing and covering problems to state the best known inapproximability results on (even highly restricted) version of Triangle Packing, 3-Set Packing, and 3-Set Covering problems (Section 4). These reductions are quite straightforward from Max-3-IS, resp. Min-3-NC and they are included as inspiration to the new reduction for 3-Dimensional Matching problem (Max-3-DM) (Section 5). APX-completeness of the problem has been well known even on instances with at most 3 occurrences of any element, but our lower bound applies to the instances with exactly 2 occurrences. We do not know about any previous hardness result on the problem with such restricted case. The best to our knowledge lower and upper approximation bounds for mentioned packing and covering problems are summarized in the following table. The upper bounds are from [5] and [6].

Problem	Param. lower bound	Approx. lower bound $(\tau_* = 6.9)$	Approx. upper bound
Max-3-DM	$1 + \frac{1}{18\tau_*+13}$	139/138	$1.5 + \varepsilon$
Max Triangle Packing	$1 + \frac{1}{18\tau_*+13}$	139/138	$1.5 + \varepsilon$
3-Set Packing	$1 + \frac{1}{18\tau_*+13}$	139/138	$1.5 + \varepsilon$
3-Set Covering	$1 + \frac{1}{18\tau_*+18}$	144/143	$1.4 + \varepsilon$

Our inapproximability result on Max-3-DM can be applied to obtain explicit lower bounds for several problems of practical interest, e.g. scheduling problems, some (even highly restricted) cases of Generalized Assignment problem, or the other more general packing problems.

2 Inapproximability of subproblems of Max-E3-Lin-2

In proving inapproximability results we produce new "hard gaps" from those already known using gap-preserving reductions and their compositions. We start with a restricted version of MAX-E3-LIN-2:

Definition 1. MAX-E3-LIN-2 *is the following optimization problem: Given a system I of linear equation over $\mathbb{Z}_2$, with exactly 3 (distinct) variables in each equation. The goal is to maximize, over all assignments ψ to the variables, the fraction of satisfied equations of I.*

We use the notation Ek-OCC-MAX-Ed-LIN-2 for the same maximization problem, where each equation has exactly d variables and each variable occurs exactly k times. If we drop an "E" than we have "at most d variables" and/or "at most k occurrences". Denote $Q(\varepsilon, k)$ the following restricted version of MAX-E3-LIN-2: *Given an instance of* Ek-OCC-MAX-E3-LIN-2. *The problem is to decide if the fraction of more than $(1 - \varepsilon)$ or less than $(\frac{1}{2} + \varepsilon)$ of all equations is satisfied by the optimal (i.e. maximizing) assignment.*

From Håstad [4] result one can prove NP-hard gap result also for instances of MAX-E3-LIN-2 where each variable appears bounded (or even constant) number of times (Theorem 1). For our applications the strengthening contained in Theorems 2 and 3 are more convenient. The proofs of Theorems 1–3 can be found in [8].

Theorem 1. (Håstad) *For every $\varepsilon \in \left(0, \frac{1}{4}\right)$ there is an integer $k_0(\varepsilon)$ such that the partial decision subproblem $Q(\varepsilon, k_0(\varepsilon))$ of* MAX-E3-LIN-2 *is NP-hard.*

Theorem 2. *For every $\varepsilon \in \left(0, \frac{1}{4}\right)$ there is a constant $k(\varepsilon)$ such that for every integer $k \geq k(\varepsilon)$ the partial decision subproblem $Q(\varepsilon, k)$ of* MAX-E3-LIN-2 *is NP-hard.*

To prove hard gap results for some problems using reduction from MAX-E3-LIN-2 it is sometimes useful, if all equations have the same right hand side. This can be easily enforced if we allow flipping some variables. The canonical gap versions $Q_i(\varepsilon, 2k)$ of MAX-E3-LIN-2 of this kind are as follows: *Given an instance of* MAX-E3-LIN-2 *such that all equations are of the form $x + y + z = i$ and each variable appears exactly k times negated and k times unnegated. The task is to decide if the fraction of more than $(1 - \varepsilon)$ or less than $(\frac{1}{2} + \varepsilon)$ of all equations is satisfied by the optimal (i.e. maximizing) assignment.* The corresponding hard-gap result for this restricted version reads as follows.

Theorem 3. *For every $\varepsilon \in \left(0, \frac{1}{4}\right)$ there is a constant $k(\varepsilon)$ such that for every integer $k \geq k(\varepsilon)$ the partial decision subproblems $Q_0(\varepsilon, 2k)$ and $Q_1(\varepsilon, 2k)$ of* MAX-E3-LIN-2 *are NP-hard.*

3 Amplifiers

In this section we describe our results about the structure and parameters of 3-regular amplifiers, that we use in our reductions.

Definition 2. *A graph $G = (V, E)$ is a $(2,3)$-graph if G contains only nodes of degree 2 and 3. We denote Contacts $= \{v \in V : \deg_G(v) = 2\}$, and Checkers $= \{v \in V : \deg_G(v) = 3\}$. Furthermore, a $(2,3)$-graph G is an* amplifier *(more precisely, it is a 3-regular amplifier for its contact nodes) if for every $A \subseteq V$: $|\mathrm{Cut}\, A| \geq |Contacts \cap A|$, or $|\mathrm{Cut}\, A| \geq |Contacts \setminus A|$, where $\mathrm{Cut}\, A = \{\{u, v\} \in E$: exactly one of nodes u and v is in $A\}$.*

An amplifier G is called a (k, τ)-amplifier if $|Contacts| = k$ and $|V| = \tau k$. We introduce the notation $\tau(G) := \frac{|V|}{|Contacts|}$ for an amplifier G. Let us denote $\tau_* = \inf\{\gamma : \tau(G) < \gamma$ for infinitely many amplifiers $G\}$.

We have studied several probabilistic models of generating $(2,3)$-graphs randomly. In such situation we need to estimate the probability that the random $(2,3)$-graph G is an amplifier. It fails to be an amplifier if and only if the system of so-called *bad sets* $\mathcal{B} := \big\{A \subseteq V : |\mathrm{Cut}\, A| < \min\{|Contacts \cap A|, |Contacts \setminus A|\}\big\}$ is nonempty. For a fixed bad set it is quite simple to estimate the probability that this candidate for a bad set doesn't occur. But the question is how to estimate the union bound over all bad sets in better way, than by adding all single probabilities. It is useful to look for a small list $\mathcal{B}_* \subseteq \mathcal{B}$, such that if $\mathcal{B} \neq \emptyset$ then $\mathcal{B}_* \neq \emptyset$ as well. In [2] the role of $\mathcal{B}_*$ play elements of $\mathcal{B}$ of the minimum size. Our analysis shows that one can produce the significantly smaller list of bad sets which is sufficient to exclude to be sure that a graph is an amplifier.

For a $(2,3)$-graph $G = (V, E)$ we define the relation $\preceq$ on the set $\mathcal{P}(V)$ of all subsets V: $A \preceq B$ iff $|\mathrm{Cut}\, A| \leq |\mathrm{Cut}\, B| - |(A \triangle B) \cap Contacts|$ whenever $A, B \subseteq V$. (Here $A \triangle B$ stands for $(A \setminus B) \cup (B \setminus A)$.) Clearly, the relation $\preceq$ is reflexive and transitive. So, $\preceq$ induces a partial order on the equivalence classes $\mathcal{P}(V) / \approx$. The equivalence relation $\approx$ can be more simply characterized by $A \approx B$ iff $A \cap Contacts = B \cap Contacts$ and $|\mathrm{Cut}\, A| = |\mathrm{Cut}\, B|$, for $A, B \subseteq V$. Moreover, for every $A \subseteq V$, $A \preceq B$ iff $V \setminus A \preceq V \setminus B$.

Using this relation one can describe the set $\mathcal{B}$ of bad sets, as

$$\mathcal{B} := \mathcal{B}(G) = \{B \subseteq V : \text{ neither } \emptyset \preceq B, \text{ nor } V \preceq B\}.$$

Clearly, for every $A, B \subseteq V$, $B \in \mathcal{B}$ and $A \preceq B$ imply $A \in \mathcal{B}$. The minimal elements of the partial order $(\mathcal{P}(V), \preceq)$ play an important role in what follows.

Further, we denote

$$\mathcal{B}_0 := \mathcal{B}_0(G) = \{B \subseteq V : B \text{ is a minimal element of } (\mathcal{B}, \preceq)\}.$$

Clearly, a set $\mathcal{B}_0$ is closed on the complementation operation $A \mapsto V \setminus A$ for any subset $A \subseteq V$.

Lemma 1. *Let G be a $(2,3)$-graph and $B \in \mathcal{B}_0(G)$ be given.*

(i) For every set $Z \subseteq B$ the inequality $2 \cdot |\mathrm{Cut}\, Z \cap \mathrm{Cut}\, B| \leq |\mathrm{Cut}\, Z| + |Z \cap Contacts|$ holds with the equality iff $B \setminus Z \approx B$. In particular, if $Z \cap Contacts \neq \emptyset$ the inequality is strict.

(ii) The set $\mathrm{Cut}\, B$ is a matching in G.

For a $(2,3)$-graph $G = (V, E)$ let $Z \subseteq V$ be given. Let $G_Z = (Z, E_Z)$ stand for the subgraph of G induced by the node set Z. To see that $|\mathrm{Cut}\, Z| +$

$|Z \cap Contacts| = 3|Z| - 2|E_Z|$, we can argue as follows:

$$|\text{Cut } Z| = \sum_{v \in Z \cap Checkers} (3 - \deg_{G_Z}(v)) + \sum_{v \in Z \cap Contacts} (2 - \deg_{G_Z}(v))$$

$$= \sum_{v \in Z} (3 - \deg_{G_Z}(v)) - |Z \cap Contacts| = 3|Z| - 2|E_Z| - |Z \cap Contacts|.$$

Given $B \in \mathcal{B}_0$, Cut B is a matching in G as follows from Lemma 1(ii). Let $Cutters(B)$ stand for the set of nodes in B adjacent to Cut B. Clearly for any $Z \subseteq B$, an edge of Cut B adjacent to $v \in Cutters(B)$ belongs to Cut Z if and only if $v \in Z$. Therefore $|\text{Cut } Z \cap \text{Cut } B| = |Z \cap Cutters(B)|$. Hence we can reformulate the first part of Lemma 1 as follows:

Lemma 2. *Let G be a $(2,3)$-graph and $B \in \mathcal{B}_0(G)$ be given. Then for every set $Z \subseteq B$ the inequality $|Z \cap Cutters(B)| \leq \frac{3}{2}|Z| - |E_Z|$ holds with the equality iff $B \setminus Z \approx B$. In particular, if $Z \cap Contacts \neq \emptyset$ the inequality is strict.*

The purpose of the following lemma is to derive some restrictions on local patterns of Cut B for a general set $B \in \mathcal{B}_0$. Given $B \in \mathcal{B}_0$, we can test it with many various $Z \subseteq B$ (typically with G_Z being a small connected graph) to obtain restrictions on possible patterns of $Cutters(B)$ in B. Some of basic results of this kind are stated in the following lemma.

Lemma 3. *Let G be a $(2,3)$ graph, $B \in \mathcal{B}_0(G)$ and $Z \subseteq B$ be given.*
   (i) If G_Z is a tree and $|Z| = 2k - 1$ ($k = 1, 2, \dots$) then $|Z \cap Cutters(B)| \leq k$.
   (ii) If G_Z is a tree and $|Z| = 2k$ then $|Z \cap Cutters(B)| \leq k + 1$. Moreover, this inequality is strict if $Z \cap Contacts \neq \emptyset$.
   (iii) If G_Z is a $(2k+1)$-cycle then $|Z \cap Cutters(B)| \leq k$.
   (iv) If G_Z is a $2k$-cycle then $|Z \cap Cutters(B)| \leq k$. Moreover, this inequality is strict if $Z \cap Contacts \neq \emptyset$.

Lemma 4. *Let $G = (V, E)$ be a $(2,3)$-graph and $B \in \mathcal{B}_0(G)$ be given.*
   (i) If $a, b \in Cutters(B)$ and $(a, b) \in E$, then $a, b \in Checkers$ and there are 2 distinct nodes a', $b' \in B \setminus Cutters(B)$ such that $(a, a') \in E$ and $(b, b') \in E$.
   (ii) If $a, c \in Cutters(B)$, $b \in B$, $(a, b) \in E$, $(b, c) \in E$, and if exactly one of nodes a, b and c belongs to Contacts, then there are 2 distinct nodes d, $e \in B \setminus Cutters(B)$, each adjacent to one of two nodes in $\{a, b, c\} \cap Checkers$.

For the purpose to provide even more restricted list we make our partial order $\preceq$ finer inside the equivalence classes $\mathcal{P}(V) / \approx$. For a $(2,3)$-graph $G = (V, E)$ let a subset F of E of "distinguished edges" be fixed. We define the following relations on the set $\mathcal{P}(V)$ of all subset V, whenever $A, B \subseteq V$: $A \overset{F}{\preceq} B$ iff either $(A \preceq B \ \& \ A \not\approx B)$ or $(A \approx B \ \& \ |F \cap \text{Cut } A| \leq |F \cap \text{Cut } B|)$; $A \overset{F}{\preceq}_* B$ iff either $(A \overset{F}{\preceq} B \ \& \ A \overset{F}{\not\approx} B)$, or $(A \overset{F}{\approx} B \ \& \ \min\{|A \cap Checkers|, |Checkers \setminus A|\} \leq \min\{|B \cap Checkers|, |Checkers \setminus B|\})$.

Denote

$$\mathcal{B}_F(G) := \{B \subseteq V : B \text{ is a minimal element of } (\mathcal{B}(G), \overset{F}{\preceq})\},$$

$$\mathcal{B}_F^*(G) := \{B \subseteq V : B \text{ is a minimal element of } (\mathcal{B}(G), \overset{F}{\preceq}_*)\}.$$

The equivalence relation $\overset{F}{\approx}$ is defined by: $A \overset{F}{\approx} B$ iff $A \overset{F}{\preceq} B$ and $B \overset{F}{\preceq} A$. Clearly $\overset{F}{\preceq}$ is a partial order on equivalence classes $\mathcal{P}(V) / \overset{F}{\approx}$. The equivalence relation $\overset{F}{\approx}$ can be also characterized by $A \overset{F}{\approx} B$ iff $A \cap Contacts = B \cap Contacts$ and $|\mathrm{Cut}\, A| = |\mathrm{Cut}\, B|$ & $|F \cap \mathrm{Cut}\, A| = |F \cap \mathrm{Cut}\, B|$.

Clearly $\mathcal{B}_F^*(G) \subseteq \mathcal{B}_F(G) \subseteq \mathcal{B}_0(G)$, and $B \in \mathcal{B}_F(G)$ iff $B \in \mathcal{B}_0(G)$ & $A \approx B$ implies $|F \cap \mathrm{Cut}\, B| \leq |F \cap \mathrm{Cut}\, A|$; $B \in \mathcal{B}_F^*(G)$ iff $B \in \mathcal{B}_F(G)$ & $A \overset{F}{\approx} B$ implies $\min\{|B \cap Checkers|, |Checkers \setminus B|\} \leq \min\{|A \cap Checkers|, |Checkers \setminus A|\}$.

Lemma 5. *Let G be a $(2,3)$-graph and $B \in \mathcal{B}_F(G)$ be given. Then for every set $Z \subseteq B$ such that $B \setminus Z \approx B$ (equivalently, $Z \subseteq B \cap Checkers$ and $|Z \cap Cutters(B)| = \frac{3}{2}|Z| - |E_Z|$) $|F \cap \mathrm{Cut}\, Z \cap \mathrm{Cut}\, B| \leq \frac{1}{2}|F \cap \mathrm{Cut}\, Z|$ holds, with the equality if and only if $B \setminus Z \overset{F}{\approx} B$.*

Lemma 6. *Let G be a $(2,3)$-graph, $B \in \mathcal{B}_F^*(G)$ and $\emptyset \neq Z \subseteq B$ such that $B \setminus Z \approx B$ and $2 \cdot |B \cap Checkers| < |Checkers| + |Z|$. Then $|F \cap \mathrm{Cut}\, Z \cap \mathrm{Cut}\, B| < \frac{1}{2} \cdot |F \cap \mathrm{Cut}\, Z|$.*

Let us consider a $(2,3)$-graph $G = (V, E)$. For $B \subseteq V$, we denote $B_{\mathrm{red}} := B \cap Checkers$. Assume further that no pair of nodes in $Contacts$ is adjacent by an edge. We convert G to a 3-regular (multi-)graph G_{red} with a node set V_{red} equals to $Checkers$. Each node $v \in Contacts$ and two edges adjacent to v in G are replaced with an edge $e(v)$ (later called a *contact* edge) that connects the pair of nodes that were adjacent to v in G. For any $A \subseteq V_{\mathrm{red}}$ let $\mathrm{Cut}_{\mathrm{red}}A$ stand for a cut of A in G_{red}, and $Cutters_{\mathrm{red}}(A)$ stand for the set of nodes of A adjacent in G_{red} to an edge of $\mathrm{Cut}_{\mathrm{red}}A$.

Lemma 7. *Let G be a $(2,3)$-graph with no edge between contact nodes, and let $B \in \mathcal{B}_0(G)$. Then $|\mathrm{Cut}\, B| = |\mathrm{Cut}_{red}(B_{red})|$, and if any pair of nodes in $Contacts$ is at least at distance 3 apart, $\mathrm{Cut}_{red}(B)$ is a matching in G_{red}.*

We elaborate in details on our general results in the concrete model of randomly generated (k, τ)-wheels, which generalizes slightly the notion of a wheel-amplifier used by Berman and Karpinski ([2]). A (k, τ)-wheel is a $(2,3)$-graph $G = (V, E)$ with $|V| = \tau k$ and $|Contacts| = k$, and with the edge set E splited into two parts E_C and E_M. E_C is an edge set of several disjoint cycles in G collectively covering V. In each cycle consecutive contacts of G are separated by a chain of several (at least 2) checkers. E_M is a perfect matching for the set of checkers. We consider here the choice $F := E_C$ for the special subset of "distinguished edges" in our amplifier analysis.

Given a bad set B, we will refer to fragments of B, the connected components of B within cycles, and to reduced fragments of B_{red}, the connected components of B_{red} within corresponding reduced cycles.

The following theorem summarizes the results from Lemmas 1–7 for (k, τ)-wheel:

Proposition 1. *Let G be a (k, τ)-wheel. Then every set $B \in \mathcal{B}_0(G)$ has the following properties:*

(i) B is a bad set, i.e. $|\text{Cut }B| < \min\{|Contacts \cap B|, |Contacts \setminus B|\}$.
(ii) $\text{Cut }B$ is a matching in G.
(iii) $|\text{Cut}_{red}(B_{red})| = |\text{Cut }B|$, and $\text{Cut}_{red}(B_{red})$ is a matching in G_{red}.
(iv) Any fragment of B contains at least 2 checkers.
(v) End nodes of any reduced fragment of B_{red} are not incident to $E_M \cap \text{Cut }B$.
(vi) Any fragment of B consisting of 3 checkers has none of its nodes incident to $E_M \cap \text{Cut }B$.
(vii) Any fragment of B consisting of 2 checkers and 1 contact has both its checkers matched with $B \setminus Cutters(B)$ nodes.
(viii) Any fragment of B consisting of 2 checkers has both its nodes matched with $B \setminus Cutters(B)$ nodes.

Every set $B \in \mathcal{B}_F(G)$ additionally has the following properties:

(ix) Any fragment of B contains at least 3 nodes.
(x) Any fragment of B consisting of 3 checkers has all its nodes matched with $B \setminus Cutters(B)$.
(xi) Any fragment of B consisting of 4 checkers has none of its nodes incident to $E_M \cap \text{Cut }B$.

All the above properties apply at the same time to B and $\tilde{B} := Checkers \setminus B$. The following is less symmetric, it says something more about the smaller of the sets B, $\tilde{B}$, if $B \in \mathcal{B}(G)_F^*$.

(xii) If $B \in \mathcal{B}(G)_F^*$ with $|B \cap Checkers| \leq \frac{1}{2}|Checkers|$, then no pair of checkers that are end nodes of (possibly distinct) fragments of B, are matched.

For purpose of the paper we can confine ourselves to the model with E_C consisting of 2 cycles C_1 and C_2. One consists of $(1 - \theta)k$ ($\theta \in (0,1)$) contacts, separated by chains of checkers of length 6, and in the second one θk contacts are separated by chains of checkers of length 5. For fixed parameters θ and k consider two cycles with contacts and checkers as above and take a random perfect matching for the set of checker nodes. Then, with high probability, the produced $(k, 7 - \theta)$-wheel will be an amplifier. More precisely, for an explicit constant $\theta_0 \in (0,1)$, for any rational $\theta \in (0, \theta_0)$, and any sufficiently large positive integer k for which θk is an even integer, $(k, 7 - \theta)$-amplifiers exist. Here τ is a rational number, $(\tau - 1)k$ is an even integer. In such model the following upper bound for τ_* can be proved

Theorem 4. $\tau_* \leq 6.9$.

The proof of this theorem is quite technical (see [8] for details and proofs of Lemmas 1–7). The further improvements of estimates on amplifier parameters of randomly generated graphs, pushing the method to its limits, is in progress.

4 Amplifier parametrized known reductions

We call HYBRID a system of linear equations over $\mathbb{Z}_2$, each equation either with 2 or with 3 variables. We are interested in hard gap results for instances

of HYBRID with exact 3 occurrences of each variable (a subproblem of E3-Occ-Max-3-Lin-2). As suggested in [2], one can produce hard gaps for such restricted instances of HYBRID by gap-preserving reduction from Max-E3-Lin-2. Our approach is simpler than in [2], since we start the reduction from the problem which is already of bounded (even constant, possibly very large) occurrence. This is a crucial point, since the number of occurrences of variables is just the value that has to be amplified using the expander or amplifier method, and in our reductions an amplifier plays a role of a constant.

Reduction from $Q(\varepsilon, k)$ to HYBRID(G). Let $\varepsilon \in \left(0, \frac{1}{4}\right)$, and $k \in \mathbb{N}$ be such that $Q(\varepsilon, k)$ is NP-hard. Now we describe a gap-preserving reduction from $Q(\varepsilon, k)$ to the corresponding gap-version of HYBRID. Assume that $G = (V, E)$ is a fixed (k, τ)-amplifier with $|Contacts| = k$ and $|V| = \tau k$. Let an instance I of $Q(\varepsilon, k)$ be given, denote by $\mathcal{V}(I)$ the set of variables in I, $m := |\mathcal{V}(I)|$. Take m disjoint copies of G, one for each variable from $\mathcal{V}(I)$. Let G_x denote a copy of G that corresponds to a variable x. The contact nodes of G_x represent k occurrences of x in equations of I. Distinct occurrences of a variable x in I are now represented by distinct contact nodes of G_x. For each equation $x + y + z = i$ of I ($i \in \{0, 1\}$) we create a hyperedge of size 3, labeled by i. A hyperedge connects a triple of contact nodes, one from each G_x, G_y and G_z. The edges inside each copy G_x are labeled by 0 and any such edge (u, v) represents the equivalence equation $u + v = 0$.

The produced instance I' of HYBRID corresponds simultaneously to a system of equations and a labeled hypergraph. Clearly, nodes correspond to variables, and labeled (hyper-)edges to equations in an obvious way. The restriction of HYBRID to these instances will be called as HYBRID(G) in what follows. The most important property of a produced instance I' is that each variable occurs exactly 3 times in equations. In particular, each contact node occurs exactly in one hyperedge. If an instance I has m variables with $|I| = \frac{mk}{3}$ equations, then I' has $m\tau k$ variables, $\frac{mk}{3}$ equations with 3 variables, and $\frac{mk}{2}(3\tau - 1)$ equations with 2 variables. Hence $|I'| = \frac{mk}{6}(9\tau - 1)$ equations in total.

Clearly, any assignment to variables from $\mathcal{V}(I)$ generates so called *standard* assignment to variables of I': the value of a variable x is assigned to all variables of G_x. To show that the optimum OPT(I') is achieved on standard assignments is easy. But for a standard assignment the number of unsatisfied equations for I' is the same as for I. Consequently, OPT(I') depends affinely on OPT(I), namely $(1 - \text{OPT}(I'))|I'| = (1 - \text{OPT}(I))|I|$. Now we see that OPT$(I) > 1 - \varepsilon$ implies OPT$(I') > 1 - \frac{2\varepsilon}{9\tau - 1}$, and OPT$(I) < \frac{1}{2} + \varepsilon$ implies OPT$(I') < \frac{9\tau - 2}{9\tau - 1} + \frac{2\varepsilon}{9\tau - 1}$. This proves that it is NP-hard to decide whether an instance of HYBRID(G) with $|I|$ equations has the maximum number of satisfied equations above $(1 - \frac{2\varepsilon}{9\tau - 1})|I|$ or below $(\frac{9\tau - 2}{9\tau - 1} + \frac{2\varepsilon}{9\tau - 1})|I|$. Hence, we have just proved the following:

Theorem 5. *Assume that $\varepsilon \in \left(0, \frac{1}{4}\right)$, let k be an integer such that $Q(\varepsilon, k)$ is NP-hard, and G be a (k, τ)-amplifier. Then it is NP-hard to decide whether an instance of HYBRID(G) with $|I|$ equations has the maximum number of satisfied equations above $(1 - \frac{2\varepsilon}{9\tau - 1})|I|$ or below $(\frac{9\tau - 2}{9\tau - 1} + \frac{2\varepsilon}{9\tau - 1})|I|$.*

Corollary 1. *It is NP-hard to approximate the solution of E3-Occ-Max-3-Lin-2 within any constant smaller than $1 + \frac{1}{9\tau_* - 2}$.*

Reductions from HYBRID(G) to other problems. We refer to [2] where Berman and Karpinski provide gadgets for reductions from HYBRID to small bounded instances of MAXIMUM INDEPENDENT SET and MINIMUM NODE COVER. We can use exactly the same gadgets in our context, but instead of their wheel-amplifier we use a general (k, τ)-amplifier. The proofs from [2] apply in our context as well.

Theorem 6. *Let $\varepsilon \in \left(0, \frac{1}{4}\right)$, $k \in \mathbb{N}$ be such that $Q(\varepsilon, k)$ is NP-hard, and τ be such that a (k, τ)-amplifier exists. It is NP-hard to decide whether an instance of* MAX-3-IS *with n nodes has the maximum size of an independent set above $\frac{18\tau + 14 - 2\varepsilon}{4(9\tau + 8)} n$, or below $\frac{18\tau + 13 + 2\varepsilon}{4(9\tau + 8)} n$. Consequently, it is NP-hard to approximate the solution of* MAX-3-IS *within any constant smaller than $1 + \frac{1}{18\tau_* + 13}$. Similarly, it is NP-hard to decide whether an instance of* MIN-3-NC *with n nodes has the minimum size of a node cover above $\frac{18\tau + 19 - 2\varepsilon}{4(9\tau + 8)} n$, or below $\frac{18\tau + 18 + 2\varepsilon}{4(9\tau + 8)} n$. Consequently, it is NP-hard to approximate the solution of* MIN-3-NC *within any constant smaller than $1 + \frac{1}{18\tau_* + 18}$. The same hard-gap and inapproximability results apply to 3-regular triangle-free graphs.*

In the following we present the inapproximability results for three similar APX-complete problems. From L-reductions used in the proofs of Max-SNP completeness (see [9] or [10]) some lower bounds can be computed but they would be worse as the lower bounds presented here.

Maximum Triangle Packing problem. A triangle packing for a graph $G = (V, E)$ is a collection $\{V_i\}$ of disjoint 3-sets of V, such that every V_i induces a 3-clique in G. The goal is to find cardinality of maximum triangle packing. The problem is APX-complete even for graphs with maximum degree 4 ([9]).

Maximum 3-Set Packing problem. Given a collection C of sets, the cardinality of each set in C is at most 3. A set packing is a collection of disjoint sets $C' \subseteq C$. The goal is to find cardinality of maximum set packing. If the number of occurrences of any element in C is bounded by a constant K, $K \geq 2$, the problem is still APX-complete ([1]).

Minimum 3-Set Covering problem. Given a collection C of subsets of a finite set S, the cardinality of each set in C is at most 3. The goal is to find cardinality of minimum subset $C' \subseteq C$ such that every element in S belongs to at least one member of C'. If the number of occurrences of any element in sets of C is bounded by a constant $K \geq 2$, the problem is still APX-complete [10].

Theorem 7. *Assume that $\varepsilon \in \left(0, \frac{1}{4}\right)$, $k \in \mathbb{N}$ is such that $Q(\varepsilon, k)$ is NP-hard, and τ is such that there is a (k, τ)-amplifier.*

(i) It is NP-hard to decide whether an instance of TRIANGLE PACKING *with n nodes has the maximum size of a triangle packing above $\frac{18\tau + 14 - 2\varepsilon}{6(9\tau + 8)} n$, or below $\frac{18\tau + 13 + 2\varepsilon}{6(9\tau + 8)} n$. Consequently, it is NP-hard to approximate the solution of* MAXIMUM TRIANGLE PACKING *problem (even on 4-regular graphs) within any constant smaller than $1 + \frac{1}{18\tau_* + 13}$.*

(ii) It is NP-hard to decide whether an instance of 3-SET PACKING *with n triples and the occurrence of each element exactly in two triples has the maximum size of a packing above $\frac{18\tau + 14 - 2\varepsilon}{4(9\tau + 8)} n$, or below $\frac{18\tau + 13 + 2\varepsilon}{4(9\tau + 8)} n$. Consequently, it*

is NP-*hard to approximate the solution of* 3-SET PACKING *with exactly two occurrences of each element within any constant smaller than* $1 + \frac{1}{18\tau_* + 13}$.

(iii) *It is* NP-*hard to decide whether an instance of* 3-SET COVERING *with n triples and the occurrence of each element exactly in two triples has the minimum size of a covering above* $\frac{18\tau + 19 - 2\varepsilon}{4(9\tau + 8)}n$, *or below* $\frac{18\tau + 18 + 2\varepsilon}{4(9\tau + 8)}n$. *Consequently, it is* NP-*hard to approximate the solution of* 3-SET COVERING *with exactly two occurrences of each element within any constant smaller than* $1 + \frac{1}{18\tau_* + 18}$.

Proof. Consider a 3-regular triangle-free graph G as an instance of MAX-3-IS from Theorem 6. (i) Take a line-graph $L(G)$ of G. Nodes of G are transformed to triangles in $L(G)$ and this is one-to-one correspondence, as G was triangle-free. Clearly, independent sets of nodes in G are in one-to-one correspondence with triangle packings in $L(G)$, so the conclusion easily follows from Theorem 6. (ii) Create an instance of 3-SET PACKING that uses for 3-sets exactly triples of edges of G adjacent to each node of G. Clearly, independent sets of nodes in G are in one-to-one correspondence with packings of triples in the corresponding instance. Now the conclusion easily follows from the hard-gap for MAX-3-IS problem. (iii) Now a graph G from Theorem 6 is viewed as an instance of MIN-3-NC. Using the same collection of 3-sets as in the part (ii) we see that node covers in G are in one-to-one correspondence with coverings by triples in the new instance. The conclusion follows from the hard-gap result for MIN-3-NC from Theorem 6.

5 New reduction for 3-Dimensional Matching

Definition and known results. Given the disjoint sets A, B, and C and a set $T \subseteq A \times B \times C$. A matching for T is a subset $T' \subseteq T$ such that no elements in T' agree in any coordinate. The goal of the MAXIMUM 3-DIMENSIONAL MATCHING problem (shortly, MAX-3-DM) is to find cardinality of a maximum matching. The problem is APX-complete even in case if the number of occurrences of any element in A, B or C is bounded by a constant K ($K \geq 3$) [9].

Recall that usually the hardness of MAX-3-DM is proved by reduction from bounded instances of MAX-3-SAT. The L reduction given in [9] implies lower bound $(1 + \varepsilon)$ for some small ε. In what follows we present the new transformation from HYBRID to edge 3-colored instances of MAX-3-IS. To the best of our knowledge we provide the first explicit lower bound on approximation of MAX-3-DM.

Idea: If we have hardness result for MAX-3-IS on 3-regular edge-3-colored graphs, it is at the same time the result for MAX-3-DM due to the following transformation. Suppose that edges of graph $G = (V, E)$ are properly colored with three colors a, b, c. Now define the sets $A = \{$all edges of color $a\}$, $B = \{$all edges of color $b\}$, $C = \{$all edges of color $c\}$ and a set $T \subseteq A \times B \times C$ as $T = \{(e_a(v), e_b(v), e_c(v))$, for all $v \in V\}$, where $e_i(v)$ denotes an edge of color i incident to the node v. It is easy to see that independent sets of nodes in G are in one-to-one correspondence with matchings of an instance obtained by the reduction above. So, the hardness result for MAX-3-DM will immediately follow from the hardness result for MAX-3-IS on edge-3-colored graphs.

Theorem 8. *Given $\varepsilon \in \left(0, \frac{1}{4}\right)$ and let k be an integer such that $Q(\varepsilon, k)$ is NP-hard. Assume τ is such that there is a (k, τ)-amplifier. Then it is NP-hard to decide whether an instance of* MAX-3-DM *with n-triples, each element occurring in exactly two triples, has the maximum size of a matching above $\frac{18\tau + 14 - 2\varepsilon}{4(9\tau + 8)}n$, or below $\frac{18\tau + 13 + 2\varepsilon}{4(9\tau + 8)}n$. Hence it is* NP-*hard to approximate* MAX-3-DM *within any constant smaller than $1 + \frac{1}{18\tau_* + 13}$ even on instances with exactly two occurrences of each element.*

Proof. Let $\varepsilon \in \left(0, \frac{1}{4}\right)$, $k \in \mathbb{N}$ be such that $Q(\varepsilon, k)$ is NP-hard, and $G = (V, E)$ be a fixed (k, τ)-amplifier. We use the same reduction of an instance of HYBRID(G) to an instance G' of MAX-3-IS as in Theorem 6. Each variable x of I is replaced with a gadget A_x. The gadget of a checker is a hexagon H_x in which nodes with labels 0 and 1 alternate. A gadget of a contact is a hexagon H_x augmented with a trapezoid T_x, a cycle of 6 nodes that shares one edge with a hexagon H_x. Again, labels 0 and 1 of nodes in those cycles alternate. If two variables x, y are connected by an equation, $x = y$, we connect their hexagons with a pair of edges, so called *connections*, to form a rectangle in which the nodes with label 0 and 1 alternate. The rectangle thus formed is a gadget of an equation with two variable (Fig. 1).
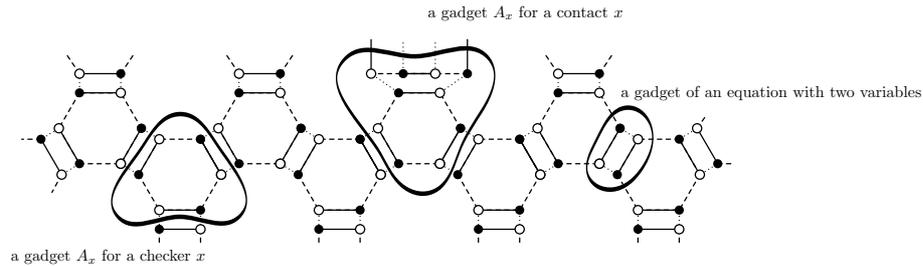

a gadget A_x for a contact x

a gadget of an equation with two variables

a gadget A_x for a checker x

Fig. 1. Example of gadgets for checkers, contacts and equations with two variables.

If three variables are connected by an equation (i.e. an hyperedge), say, $x + y + z = 0$, the trapezoids T_x, T_y and T_z (6-cycles) are coupled with the set S_{xyz} of four *special* nodes (Fig. 2).

It is easy to see that G' is 3-edge-colored. As follows from Fig. 2, the edges of the equation gadget can be colored in such way that all edges adjacent to contacts are colored by one fixed color, say A (dotted lines). All *connections* are of the same color A, which alternates on rectangles with color B (full lines), see Fig. 1. The hard gap of MAX-3-IS from Theorem 6 implies the hard gap of MAX-3-DM.

Conclusion. There is still substantial gap between the lower and upper approximation bounds for small occurrence combinatorial optimization problems. The method of parametrized amplifiers shows better the quality of used reductions and the possibilities for further improvement of lower bounds. But it is quite possible that the upper bounds can be improved more significantly.
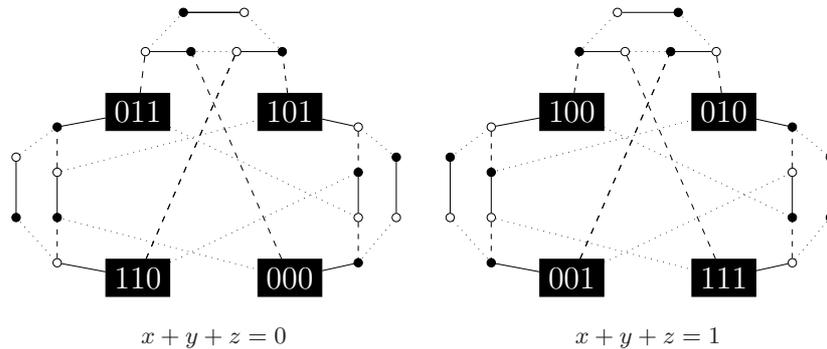
Fig. 2. Equation gadgets with three variables.

References

1. P. Berman and T. Fujito: *Approximating independent sets in degree* 3 *graphs*, Proc. 4th Workshop on Algorithms and Data Structures, 1995, Springer-Verlag, Berlin, LNCS **955**, 449–460.
2. P. Berman and M. Karpinski: *On some tighter inapproximability results, further improvements*, ECCC, Report No. **65**, 1998.
3. L. Engebretsen and M. Karpinski: *Approximation hardness of TSP with bounded metrics*, Proceedings of 28th ICALP, 2001, LNCS **2076**, 201–212.
4. J. Håstad: *Some optimal inapproximability results*, Journal of ACM **48** (2001), 798–859.
5. M. M. Halldórsson: *Approximating k-set cover and complementary graph coloring*, Proc. 5th International Conference on Integer Programming and Combinatorial Optimization, 1996, Springer-Verlag, Berlin, LNCS **1084**, 118–131.
6. C. A. J. Hurkens and A. Schrijver: *On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems*, SIAM J. Discrete Mathematics **2** (1989), 68–72.
7. M. Chlebík and J. Chlebíková: *Approximation Hardness of the Steiner Tree Problem on Graphs*, Proceedings of the 8th Scandinavian Workshop on Algorithm Theory, SWAT 2002, Springer, LNCS **2368**, 170–179.
8. M. Chlebík and J. Chlebíková: *Approximation Hardness for Small Occurrence Instances of NP-Hard Problems*, ECCC, Report No. **73**, 2002.
9. V. Kann: *Maximum bounded* 3-*dimensional matching is MAX SNP-complete*, Information Processing Letters **37** (1991), 27–35.
10. C. H. Papadimitriou and M. Yannakakis: *Optimization, approximation, and complexity classes*, J. Computer and System Sciences **43** (1991), 425–440.
11. C. H. Papadimitriou and S. Vempala: *On the Approximability of the Traveling Salesman Problem*, Proceedings of the 32nd ACM Symposium on the theory of computing, Portland, 2000.
12. M. Thimm: *On the Approximability of the Steiner Tree Problem*, Proceedings of the 26th International Symposium, MFCS 2001, Springer, LNCS **2136**, 678–689.